Look for the final version in COGNITION (2009)

When Some is Actually All: Scalar Inferences in Face-Threatening Contexts

Jean-François Bonnefon
CNRS and Université de Toulouse

Aidan Feeney
Queen's University Belfast

Gaëlle Villejoubert
Université de Toulouse

Accounts of the scalar inference from 'some *X*-ed' to 'not all *X*-ed' are central to the debate between contemporary theories of conversational pragmatics. An important contribution to this debate is to identify contexts that decrease the endorsement rate of the inference. We suggest that the inference is endorsed less often in face-threatening contexts, i.e., when *X* implies a loss of face for the listener. This claim is successfully tested in Experiment 1. Experiment 2 rules out a possible confound between face-threatening contexts and lower-bound contexts. Experiment 3 shows that while saying 'some *X*-ed' when one knew for a fact that all *X*-ed is always perceived as an underinformative utterance, it is also seen as a nice and polite thing to do when *X* threatens the face of the listener. These findings are considered from the perspective of Relevance Theory as well as that of the Generalized Conversational Inference approach.

Introduction

Individuals have a well-documented tendency to interpret the existential quantifier 'some' as conveying the negation of the universal quantifier 'all.'

(1) a. Some students stayed on campus this weekend;
b. Not all students stayed on campus this weekend.

Most adult speakers of English would assume the assertion of (1a) to imply that the speaker believes (1b). More generally, any assertion of the form 'some people *X*-ed' or 'some *F*s are *G*s' is usually taken to mean that not all people *X*-ed, or that not all *F*s are *G*s, respectively.

Grice (1989) suggested that the derivation of (1b) from (1a) reflects the general assumption that speakers abide by the maxim of Quantity, that is, that they use the maximally informative wording that is allowed by their epistemic state. If one knows that all students stayed on campus, then one ought to say so, rather than using the less informative (yet logically consistent) wording (1a). Horn (1984) identified this derivation as an instance of the more general class of *scalar inferences:* The assertion of a sentence containing an item belonging to an ordered informativeness scale such as <some, all> or <possible, certain> conversationally implies that the speaker was not in a position to use a stronger item from that scale.

Behavioural studies of the scalar inference from 'some' to 'not all' have been conducted with children (Feeney, Scrafton, Duckworth, & Handley, 2004; Noveck, 2001; Papafragou & Musolino, 2003) and adults (Bezuidenhout & Cutting, 2002; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Feeney et al., 2004; Huang & Snedeker, in press). Most of this empirical work strived to resolve the dispute about the default character of the inference (see Noveck & Reboul, 2008, for a broader review). One type of theoretical account (e.g., Chierchia, 2004; Levinson, 2000) endorses the *generalized* view that the inference is an implicature that is derived by default, but then cancelled in some contexts.[1] Another account (e.g., Carston, 1998; Sperber & Wilson, 1986/1995) endorses the *particular* view that the inference is an explicature that is not derived by default, but invited purely as a function of context.

From an empirical perspective, a critical difference between the two accounts is the time and effort required by the narrowed interpretation of 'some,' i.e., *some but not all.* Because this interpretation reflects the derivation of the scalar

Jean-François Bonnefon and Gaëlle Villejoubert, Laboratoire Cognition, Langues, Langage et Ergonomie (CNRS, UTM, EPHE), Toulouse, France. Aidan Feeney, School of Psychology, Queen's University Belfast, Northern Ireland. We thank Bénédicte Bès, Wim De Neys, and Jean-Baptiste Van der Henst for their thougthful comments. This research was supported by grant ANR-07-JCJC-0065-01 from the Agence Nationale de la Recherche, and by grants PN 08.010 (awarded by the British Council) and 18223VF (awarded by the Ministère des Affaires étrangères et Européennes) under the Alliance Franco-British Programme 2008.

Correspondence concerning this article should be sent to Jean-François Bonnefon, Laboratoire Cognition, Langues, Langage et Ergonomie, Maison de la recherche, 5 allées A. Machado, 31058 Toulouse Cedex 9, France. E-mail: bonnefon@univ-tlse2.fr.

[1] Although we group them for convenience, there are important differences between the accounts offered by Chierchia and Levinson. In particular, although Chierchia holds that scalar implicatures usually occur by default, because he argues that they are grammar-driven, he predicts that they only occur automatically in upward entailing contexts. In downward entailing contexts they are said to be effortful and to occur less frequently.

inference, it should be comparatively quick and easy according to the generalized view, whereas it should be comparatively slow and effortful according to the particular view. Current evidence tends to favour the particular account. For example, adults tend to take longer to reach the narrowed interpretation of 'some' (Bott & Noveck, 2004; Noveck & Posada, 2003). Furthermore, the narrowed interpretation is less frequent when participants are instructed to respond quickly (Bott & Noveck, 2004), or when they must simultaneously carry out a secondary task (De Neys & Schaeken, 2007).

Breheny et al. (2006) proposed another technique to test between the two accounts, by capitalising on the critical role that both accounts give to the context of the utterance. The generalised account assumes that contextual processing is involved in cancelling the narrowed interpretation, whereas the particular account assumes that contextual processing is involved in reaching the narrowed interpretation. As a consequence, the two accounts make different predictions about the time required to interpret 'some' utterances when context is manipulated so as to make the narrowed interpretation appropriate or not. The particularised account predicts that interpretation should be faster when the narrowed interpretation is contextually inappropriate, whereas the generalised account makes the opposite prediction.

Such contextual manipulations certainly allow powerful empirical tests of pragmatic theories; but they also require the identification of contexts that make the scalar inference inappropriate. There is, however, surprisingly little work on this issue, which has only addressed so far the case of *lower bound contexts.* Lower-bound contexts are such that it is enough to know that some *X*-ed, and knowing that all *X*-ed has no additional informational value.

For example, imagine that a company considers building a factory near a lake, but that legislation prohibits building factories anywhere near the natural habitat of any endangered species. An environmental expert tells the company lawyer that:

(2) Some birds at that lake belong to endangered species.

In this situation, knowing that all birds at the lake belong to endangered species is of no further import than knowing that some birds do. The bottom line, in both cases, is that the factory cannot be built there. In such a context, (2) does not seem to warrant the inference that not all the birds at the lake belong to endangered species.

Lower-bound contexts were empirically investigated by Breheny et al. (2006), who found that they decreased the processing time of scalar terms, as predicted by the particular account. This investigation was an important first step in a new direction: testing different accounts of scalars in contexts wherein the scalar inference is inappropriate. To move the field further along this path, however, it is necessary to identify new contexts wherein the scalar inference is inappropriate, beyond the only known case of lower-bound contexts. Our objective in this article is to introduce one new such class of contexts. In three experiments, we will show that *face-threatening contexts* make the narrowed interpretation of 'some' less appropriate, and that this phenomenon is unrelated to lower-boundedness.

The data we report in this article are not in any way meant to definitively prove or disprove the particular or the generalised views; rather, they are meant to move the field forward by broadening the theoretical and empirical scope of this debate. At the empirical level, finding a context that makes scalar inferences inappropriate amounts to opening a new avenue of research about scalars. Not only does it offer a solution to researchers who already wish to manipulate context so as to make the narrowed interpretation inappropriate; but it also creates the need to generalise to this new context previous findings about the time course, the development, and the processing requirements of the scalar inference. Finally, identifying a new context that makes the scalar inference inappropriate has, on its own, theoretical implications for current theories of scalars. Accordingly, we will consider in the General Discusion section whether (and at what cost) current accounts can be reconciled with our own findings. In that discussion, we will give special attention to Relevance Theory, the main representant of the particular view on scalars, and probably the most successful account of scalars so far.

Face-Threatening Contexts

Consider the following exchange:

(3) a. What impression did I make during dinner?
b. Some thought you drank too much.

We argue that (3b) leaves open the possibility that all guests at the dinner thought so, as well as the possibility that the speaker knows for a fact that all guests at the dinner thought so. More generally, we suggest that the inference from 'some *X*-ed' to 'not all *X*-ed' is made less available when *X* threatens the face of the listener. Brown and Levinson (1978/1987) define *face* as a sense of positive identity and public self-esteem that all humans project and are motivated to support in social interactions. Many actions, called face-threatening acts, can induce a loss of face for the speaker or for the listener (e.g., apologizing to other people, criticizing them). Performing such an action often requires a linguistic strategy that mitigates the face threat. One of these strategies is to use probability expressions, not to convey uncertainty, but rather to sugar-coat a face-threatening act (Bonnefon & Villejoubert, 2005, 2006):

(4) a. My decision was possibly unfair;
b. You are probably going to become deaf soon.

When they read statements such as (4a) or (4b), individuals interpret the expressions 'possibly' and 'probably' as denoting very high probabilities, because they construe them as politeness markers, i.e., as a way to preserve the faces of those involved in the conversational exchange. For example, in (4a), using 'possibly' may either save the face of the listener, by not directly contradicting her opinion that the decision was unfair, or that of the speaker, by not directly admitting that the decision was unfair. In such cases, individuals

disregard the scalar inference from 'possibly' or 'probably' to 'not certainly' out of politeness concerns: In particular, when *X* in 'possibly *X*' poses a face-threat to the listener, there is a powerful reason why, contra the maxim of Quantity, the speaker would use the term 'possibly' although she was in an epistemic position to use the term 'certainly.'

Similarly, we suggest that individuals are likely to construe the quantifier 'some' in 'some *X*-ed' as a polite device meant not to hurt the feelings of the listener, when *X*-ing is something that threatens the face of the listener. In this situation, there is again a powerful reason why, contra the maxim of Quantity, the speaker would use the term 'some' although she was in an epistemic position to use the term 'all.' If our analysis is correct, individuals should be less likely to consider that 'some X-ed' means 'not all X-ed' when X threatens the face of the listener. This prediction is tested in Experiment 1.

Experiment 1

Method

Participants were 53 undergraduate students at the University of Toulouse (9 men, 44 women, mean age 21.5, *SD* = 1.6). Each participant read two stories (Poem and Recipe), one in the face-boost condition, one in the face-threat condition (the two stories were rotated across condition, and the order of the conditions was counterbalanced across questionnaires). The face-threat version of the Poem scenario read:

> Imagine that you have joined a poetry club, which consists of 5 members in addition to you. Each week, one member writes a poem, and the 5 other members discuss the poem in the absence of its author. This week, it is your turn to write a poem and to let others discuss it. After the discussion, one fellow member confides to you that 'Some people hated your poem.'

In the face-boost version, 'Some people *hated* your poem' was replaced with 'Some people *loved* your poem.' The Recipe story was very similar to the Poem story, except that it involved a cooking club rather than a poetry club. After reading each scenario, participants answered the following Yes/No question: 'From what this fellow member told you, do you think it is possible that everyone hated [loved] your poem [recipe]?'

Results and Discussion

Overall, 83% of participants answered 'No' when asked if it was possible that everyone loved their poem/recipe when told that 'some' did. This result reflects the classic scalar inference from 'some' to 'not all.' However, a reliably lower 58% of participants answered 'No' when asked if it was possible that everyone hated their poem/recipe when told that 'some' did (Wilcoxon, $Z = 3.36$, $p < .001$, two-tailed, $h = 0.36$). To be precise, 30 out of 53 participants derived the scalar inference in both contexts, 8 did not derive it in either context, 14 derived it only in the face-boost context, and 1 derived it only in the face-threat context.

This result suggests that people's tendency to draw the scalar inference from 'some *X*-ed' to 'not all *X*-ed' decreases when *X* threatens the face of the listener. We were concerned, however, that our face-boost and face-threat conditions may correspond to an upper-bound and a lower bound context, respectively. Remember that a lower bound context is one where it is important to know that some *X*-ed, and where knowing that all *X*-ed has no additional informational value. It might be the case that love evokes upper-bound context whilst hate evokes lower-bound contexts.

More precisely, when appraising the quality of a poem or a recipe, the fact that all loved it undoubtedly carries more information than the fact that some did. This unanimity makes us more confident that the poem or recipe was indeed truly good. But is knowing that all hated our poem really more useful to us than knowing that some did? The fact that some hated our poem is likely information enough for many of us to bury it for life and not to read it again to any other audience. Learning that everyone actually hated the poem does not affect the decision we are already likely to make from the fact that some hated it. If this is the case, then sentences such as 'some hated your poem' would evoke lower-bound contexts and counter the scalar inference to 'not all,' independently of any politeness considerations. To make a stronger case for our hypothesis, we need to rule out this alternative explanation. This can be done by manipulating the target of the face-threat, making it either the listener or another person, one that neither the speaker nor the listener care for. Imagine for example that you had a chili in a restaurant and are considering recommending the place to some friends. Someone who was there with you tells you that 'some people hated their chili.' Compare this with the situation where you cooked a chili for some friends, and are considering cooking the same chili for other guests. One of the friends who was there tells you that 'some people hated your chili.'

Our politeness analysis predicts that the inference is unlikely to be derived in the second case and in the second case only. Only in the second case will the use of 'some' be construed as a politeness strategy, because neither the speaker nor the listener care for the person whose face is threatened in the first case. The alternative explanation in terms of a lower-bound context predicts that the inference is unlikely to be derived in both cases. The lower-bound context explanation assumes that it is enough to know that some guests hated the chili to decide not to cook it again (and that knowing that all guests hated the chili is not going to change that decision anyway). It is also enough to know that some people hated their chili to decide not to recommend a chili place to some friends (and, again, knowing that everybody hated their chili is not going to change that decision anyway). Experiment 2 was conducted to compare the two explanations. Furthermore, Experiment 2 was conducted to extend our initial finding to new scenarios and new measures of the interpretation of 'some.'

Experiment 2

Method

Participants were 120 undergraduate students at the University of Durham (76 men, 40 women, 4 of undisclosed gender; mean age 19.1, *SD* = 1.1). They were randomly assigned to one of the *Target* groups (Listener vs. Other). Each participant read two stories (Trip and Chili), one in the face-threat condition, the other in the face-boost condition (the two stories were rotated across condition, and the order of the conditions was counterbalanced across questionnaires). In the Target-Other group, the face-boost version of the Trip scenario read:

> Imagine you went on a group trip. You are discussing the trip with Alice, who was in the group. There were 6 other people who went on this trip. You are considering whether to recommend the trip to some friends. Hearing this, Alice tells you that 'Some people loved the way the trip was organised.'

In the Target-Listener group, the face-boost version read:

> Imagine you organised a group trip. You are discussing the trip with Alice, who was in the group. There were 6 other people who went on this trip. You are considering whether to organise this same trip again next year. Hearing this, Alice tells you that 'Some people loved the way you organised the trip.'

In the face-threat condition, the words 'some people *loved*…' were replaced with 'some people *hated*…' The Chili scenario was similar to the Trip scenario, except that it involved recommending a chili restaurant to some friends vs. cooking the same chili again for other guests.

After reading each scenario, participants rated, on a 10-point scale anchored at *totally unlikely* and *totally likely*, how likely it was that the speaker would use the word 'some' if she knew that the number of people who loved/hated the chili (or the way the trip was organised) was in fact 1. The question was then repeated for all numbers up to 6.[2]

Note that this procedure makes it explicit that the speaker is omniscient about the real number of persons who loved/hated the trip or the Chili, whereas the omniscience of the speaker was not explicitly asserted in Experiment 1. This procedure elicited the *fuzzy membership function* (Zadeh, 1965) of the concept of 'some' as a function of our experimental variables. In the context of this experiment, the membership function assigns a number (1–10) to each quantity (from *1 out of 6* to *6 out of 6*) that represents its degree of membership in the concept defined by 'some.' A membership of 1 denotes a quantity that is absolutely not in the concept, and a membership of 10 denotes a quantity that is a perfect exemplar of the concept. Other values represent intermediate degrees of membership. The membership function method provides a subtle and rich representation of the meaning of vague expressions, and has been carefully validated in many studies (see for reviews Budescu & Wallsten, 1995; Karelitz & Budescu, 2004; but see Williamson, 1994 for an argument that fuzzy logic does not provide an adequate analysis of vague expressions.)

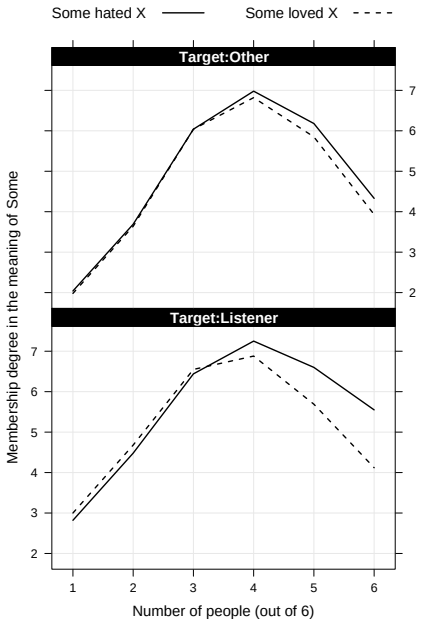

Figure 1. Membership functions of 'some people hated *X*' and 'some people loved *X*' where *X* is either something the listener did or something that some unimportant stranger did.

Results

Membership functions were computed by averaging membership judgments across participants. Figure 1 depicts the function values depending on the Target group (Listener vs. Other) and the polarity of *X* in 'some people *X*-ed' (face-boost vs. face-threat). Two pieces of information are of special interest to our purpose: the membership of the number 6 (i.e., to what extent does 'all' enter in the concept of 'some'?), and the peak of the function (i.e., what quantity is the best exemplar of the concept of 'some'?). The peak is computed by averaging for each participant the quantities that received the highest membership ratings, then averaging across participants the values so obtained.[3]

Membership of 6. As expected, there was no detectable main effect of face-boost vs. face threat ($F(1, 118) = 1.1$, $p = .30$, $\eta^2 < .01$), nor of target group ($F(1, 118) = 1.5$, $p = .22$, $\eta^2 = .01$). However, as expected, an interaction

[2] As a subsidiary question, participants also indicated whether the speaker used the word 'some' because she knew a small number did, or because she knew that a large number did, but she thought it would be more polite to say 'some.' Many participants failed to answer that question, and the number of missing answers made it impossible to analyse responses.

[3] Note that the apex of the curve is not a visual representation of the peak of the membership function. The apex of the curve is *the response option* that received, on average, the highest rating. The peak of the membership function is *the average of the response options* that received the highest rating. The graphical representation is very useful to assess trends in membership, but the value of the peak cannot be read off the graph.

effect was detected, $F(1, 118) = 3.9$, $p = .05$, $\eta^2 = .03$. This interaction reflected the fact that face-threatening contexts only had a detectable effect in the Target-Listener group. In this group, the membership of 6 was only 4.1 ($SD = 3.5$) in the face-boost condition, but went up to 5.6 ($SD = 3.7$) in the face-threat condition, $F(1, 64) = 12.8$, $p < .001$, $\eta^2 = .17$. In contrast, in the Target-Other group, the membership of 6 was 3.9 ($SD = 3.2$) in the positive condition, and 4.3 ($SD = 3.3$) in the negative condition. This difference did not achieve statistical significance, $F(1, 54) = 1.7$, $p = .20$, $\eta^2 = .03$.

Another relevant statistic is the proportion of participants who circled '10' to indicate the membership degree of the number 6. This proportion was 10% in the Target-Other group, in each face condition. But in the Target-Listener group, this proportion went from 17% in the face-boost condition to 29% in the face-threat condition.

Peak. As expected again, there was no detectable main effect of the face condition ($F(1, 118) < 0.01$, $p = .96$, $\eta^2 < .01$) or target group ($F(1, 118) = 1.1$, $p = .29$, $\eta^2 = .01$. However, the interaction was marginally significant, $F(1, 118) = 3.1$, $p = .08$, $\eta^2 = .03$. This comes close enough to significance to justify running separate analyzes in the two target groups. These analyses reveal that face-threatening contexts only have a detectable effect in the Target-Listener group. In the Target-Other group, the peak was 4.3 ($SD = 1.1$) in the face-boost condition, and 4.2 ($SD = 1.1$) in the face-threat condition. This difference did not achieve statistical significance, $F(1, 54) < 0.01$, $p = .95$, $\eta^2 < .01$. In contrast, in the Target-Listener group, the peak was 3.8 ($SD = 1.4$) in the face-boost condition, but went up to 4.2 ($SD = 1.3$) in the face-threat condition, $F(1, 64) = 5.3$, $p < .025$, $\eta^2 = .08$.

Discussion

Results of Experiment 2 further support our hypothesis that the inference from 'some *X*-ed' to 'not all *X*-ed' is less likely to be derived when *X* threatens the face of the listener. When this is the case, 'some' is understood to mean a larger proportion—but also to leave open the possibility that the speaker knew everyone *X*-ed. For example, about one-third of participants responded that it was *absolutely likely* that a speaker would say 'some people hated your chili' when she knew for a fact that everyone hated the chili cooked by the listener. Results of Experiment 1 are thus replicated in Experiment 2, using different materials and a more sophisticated measure of the interpretation of 'some.'

Results rule out a lower-bound context explanation of the findings. Such an explanation assumes that, e.g., the fact that some hated the chili is decisive enough, and that no useful information is gained by knowing that everyone hated the chili. If this explanation was correct, then the inference would also be less likely to be derived in the situation where the listener seeks advice about whether to recommend a chili place to some friends. But what we observed was the critical interaction we expected: The inference is derived less often only in situations where *X* is a face-threat to the listener. When seeking advice about whether to recommend a chili place to friends, the assertion that 'some people hated their chili' does seem to invite the inference that not all did.

Experiment 3

So far, we have shown that the interpretation of 'some X-ed' depends on whether X threatens the face of the listener. If it does, then 'some' is less likely to be interpreted as implying 'not all.' We predicted this effect from a face-management analysis of statements including 'some.' We argued that when X-ing is face-threatening, listeners may consider that 'some X-ed' was meant to stand for 'all X-ed,' out of kindness and consideration for their feelings.

In our third experiment, we test the assumption that listeners acknowledge this intention, and recognise the fact that while using 'some' when one knew *all* is a breach of the maxim of Quantity (or of the principle of relevance), it is also a nice and considerate thing to do in face-threatening contexts.

Whether *X* in 'some *X*-ed' is a face-boost or a face-threat to the listener, we expect that individuals find it dishonest and inaccurate to use 'some' in situations where the speaker knew for a fact that all *X*-ed. In contrast, we expect a critical interaction effect regarding the extent to which individuals find it nice and considerate to use 'some.' More precisely, we expect that using 'some' when one knew enough to use 'all' is judged as kind and considerate in face-threatening contexts, but unkind and inconsiderate in face-boosting contexts.

Method

Participants were 50 undergraduate students at the University of Durham (8 men, 42 women, mean age 20.4, $SD = 2.4$). They read four stories wherein *X* in 'some people *X*-ed' was either a face-boost ('some people loved') or a face-threat ('some people hated'), and where the proportion of people who *X*-ed, as known by the speaker at the time of the utterance, was either 3/6 or 6/6. For example, in the face-boost, 6/6 condition, the Speech scenario read:

> Imagine you gave a speech at a small political rally. You are discussing your speech with Denise, who was in the audience. There were 6 other people in the audience. You are considering whether to give this same speech to another audience. Hearing this, Denise tells you: 'Some people loved your speech.' Denise knows that 6 people out of 6 loved your speech.

Note that, just as in Experiment 2, this procedure makes it explicit that the speaker is omniscient about how many people loved/hated the speech. The other stories (Chili, Poem, Trip) were introduced in the Method sections of Experiments 1 and 2. After reading each story, participants rated how *accurate, considerate, honest,* and *nice* it was of the speaker to use the word 'some' in that context, using four separate 10-point scales. There were four different versions of the booklet. In each version, the experimental conditions were assigned to a different story, and the conditions were presented in a different random order.

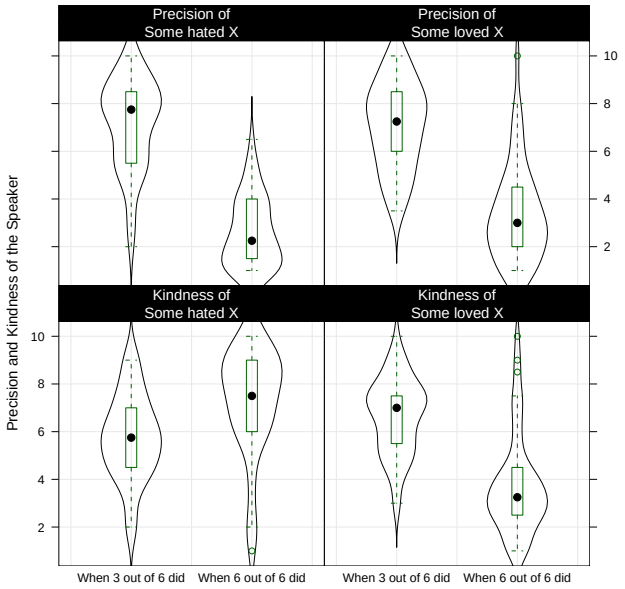

Figure 2. Precision and Kindness attributed to the speaker asserting 'some people *X*-ed,' as a function of whether *X* is a face-boost (*X* = loved) or a face-threat (*X* = hated) to the listener, and of the true proportion of people who *X*-ed (3/6 vs. 6/6).

Results

An index of *Kindness* was computed by averaging the ratings for 'considerate' and 'nice' (between which the correlation was higher than .80), and an index of *Precision* was computed by averaging the ratings for 'accurate' and 'honest' (between which the correlation was higher than .80). Figure 2 depicts the distribution of the two indices as a function of our experimental variables.

As expected, Precision was rated much higher in the 3/6 condition ($M = 7.1$, $SD = 1.7$) than in the 6/6 condition ($M = 3.0$, $SD = 1.4$), $F(1, 49) = 182.6$, $p < .001$, $\eta^2 = .79$. This was true in both the face-threat and face boost conditions, although the effect was slightly stronger in the former, as detected by the interaction effect ($F(1, 49) = 7.5$, $p = .008$, $\eta^2 = .13$).

In contrast, Kindness ratings clearly showed the expected interaction effect ($F(1, 49) = 94.0$, $p < .001$, $\eta^2 = .66$). In the face-threat condition, Kindness ratings were higher in the 6/6 condition ($M = 7.3$, $SD = 2.2$ for 6/6 vs. $M = 5.7$, $SD = 1.8$ for 3/6), but the opposite was true in the face-boost condition ($M = 3.7$, $SD = 2.1$ for 6/6 vs. $M = 6.7$, $SD = 1.6$ for 3/6).

Besides the interaction effect, two main effects were detected. Both effects appear to be due to the extremely low kindness ratings given in the face-boost, 6/6 story. For this reason, and because a very large interaction effect was detected, we are inclined not to over-interpret these two main effects, which we nevertheless report. Kindness ratings were higher in the face-threat condition ($M = 6.5$, $SD = 1.7$ vs. $M = 5.5$, $SD = 1.4$), $F(1, 49) = 20.2$, $p < .001$, $\eta^2 = .29$; and they were higher in the 3/6 condition ($M = 6.2$, $SD = 1.4$ vs. $M = 5.5$, $SD = 1.5$), $F(1, 49) = 7.8$, $p = .007$, $\eta^2 = .14$.

Discussion

Results of Experiment 3 show that saying 'some *X*-ed' when one knew enough to say 'all *X*-ed' is always perceived as inaccurate and even dishonest wording, whether or not *X* threatens the face of the listener. However, using 'some' in such a situation is also perceived as a nice and considerate thing to do when *X* threatens the face of the listener.

Participants thus appeared to acknowledge the conflict that arises between being a good Gricean speaker and being a nice person. They appeared to realise that when *X* threatens the face of the listener and one knows that everyone *X*-ed, saying that 'some *X*-ed' is a breach of Gricean or relevance-theoretic expectations, witness their responses to the Accuracy and Honesty questions. But they also appeared to realise that this linguistic formulation was the act of a kind speaker, witness their responses to the Niceness and Considerateness questions.

General Discussion

The scalar inference from 'some *X*-ed' to 'not all *X*-ed' is so intuitively compelling that some theorists (but not all) have argued that it constitutes a default reaction, which is then cancelled in some contexts (Levinson, 2000). Other theorists have argued that this inference is purely a matter of context, just like any other (particularized) conversational inference, whose derivation is a matter of applying general principles of communication, such as the maxim of Quantity (Grice, 1989), or the principle of relevance (Sperber & Wilson, 1986/1995).

Both accounts assume that the scalar inference should be available in some contexts but not in others, but make different predictions as to how such contextual aspects are processed. An important contribution to future experimental tests of the theories then consists of identifying contextual manipulations that can turn the scalar inference from available to unavailable. Very few such manipulations, though, have been proposed so far. In this article, we have identified a new and simple contextual manipulation that affects the availability of the scalar inference: When *X* in 'some *X*-ed' threatens the face of the listener, individuals are less likely to infer that the speaker meant or knew that not all *X*-ed—and this is because they consider the possibility that the speaker might want to be nice more than to be precise.[4] Although,

[4] Note however that the speaker cannot be formally accused of lying. If one believes that all *X*-ed, then one also believes that some *X*-ed, and one cannot be accused of lying by saying that some *X*-ed. We may talk about this strategy as 'sugar-coating,' but we must not confuse this kind of polite sugar-coating (Bonnefon & Villejoubert, 2006; Demeure, Bonnefon, & Raufaste, 2008, 2009) with another kind of face-saving move, informally known as the *white lie,* wherein what is asserted is downright false. These face-saving moves should not be confused either with the *self-serving use* of ambiguous language. It is quite possible, though, that contextual manipulations based on self-serving uses of language may yield

as previously stated, the purpose of this research was not to disprove either the generalized or the contextual account of scalar inferences, these findings already have significant implications for both accounts, which we now consider in turn.

Implications for the Generalized Account

The generalized account rests on the plausible assumption that deriving the scalar inference by default (and then cancelling it, in the rare cases when it is not appropriate) allows for quick and efficient communication. As noted by Noveck and Sperber (2007), though, this gain in efficiency decreases with the frequency of the situations in which the default inferences must be cancelled. If this frequency reaches some critical level, it can become less efficient to derive the inference by default.

Explaining our results within the generalized account would amount to considering that the default inference from 'some' is cancelled in face-threatening contexts. The problem with this solution is that these contexts are likely to be far from rare. To get a rough order of magnitude for their frequency, consider that Youmans (2001) analyzed taped American English conversations (23,000 words, including about 500 probability terms) and reported that probability terms were used as face-management devices (as in 'My decision was possibly unfair') at the rate of 4 out of 10 occurrences. If the polite use of 'some' is anywhere near as frequent, then the rationale for turning the scalar inference from 'some' into a default is called into question, because too many exceptions to the default quickly make it counter-productive.

Implications for Relevance Theory

From a relevance-theoretic point of view (Sperber & Wilson, 1986/1995), the scalar inference from 'some' to *some but not all* is the result of a presumption of optimal relevance. According to the original formulation of the presumption of optimal relevance, an optimally relevant act of communication (a) achieves cognitive effects that are large enough to make it worth processing; and (b) is the most relevant one that the communicator could have used to produce these effects (e.g., they cannot be achieved by another act of communication that would require less processing effort). To interpret a statement, listeners are assumed to follow a path of least effort, and to stop as soon as an interpretation of the statement meets their expectations of optimal relevance.

Outside lower-bound contexts, the unenriched interpretation *some and possibily all X-ed* does not yield enough cognitive effects to justify the effort of processing 'some.' Listeners then move on to the enriched interpretation *some but not all X-ed.* This interpretation yields enough effects to justify the additional processing effort; and these effects could not be achieved by a simpler statement than 'some *X*-ed.' Listeners thus settle on this interpretation that meets their expectation of optimal relevance.

The problem is that this explanation should apply the same whether *X*-ing is a face-boost or a face-threat to the listener. In contrast, our results clearly show that the scalar inference is less frequent when *X*-ing is a face-threat to the listener; and that listeners may even consider the interpretation *all X-ed* in that situation. Remember, for example, that 29% of participants in Experiment 2 gave a rating of 10 (out of 10) to the likelihood of saying "some people hated your chili" when knowing for a fact that everyone hated the chili. Relevance theorists may consider two ways out of this difficulty, which we will discuss in turn.

Redefining cognitive effects. It could be tempting to handle face-threatening contexts by directly factoring face-management into the computation of cognitive effects. This would amount to considering that, ceteris paribus, an utterance that preserves the face of the listener has greater cognitive effects than an utterance that hurts the feelings of the listener. With this assumption, listeners would stop at the unenriched interpretation *some and possibly all hated your poem,* because this unenriched interpretation already yields sufficient 'effects' (redefined) to justify the effort of processing it; thus the non-derivation of the scalar inference.

This solution is intuitively compelling, but its far-reaching implications are hard to calculate, for it would involve a major redefinition of a core notion within Relevance Theory, that of a cognitive effect. Up to now, cognitive effects have been diversely but consistently defined in terms of their epistemic impact on the listener (the extent to which they increase the listener's knowledge), to the exclusion of their affective impact (the way they make the listener feel). As a matter of fact, Haugh (2003) argues that this focus on epistemic rather than affective effects seriously limits the ability of Relevance Theory to account for politeness phenomena. It remains to be seen whether Relevance Theory will undergo this major transformation in order to account for the effect of face-threatening contexts, but this evolution seems unlikely. Indeed, Relevance Theory does not seem to be moving towards extending the definition of cognitive effects, but, as we shall see, rather towards extending the definition of optimal relevance. We now consider how this evolution may account for the effect of face-threatening contexts.

Redefining optimal relevance. In the postface to the 1995 edition of *Relevance*, Sperber and Wilson introduced a revised formulation of the presumption of optimal relevance. Just as the original formulation, this revised version has two clauses. The first clause is kept unchanged: an optimally relevant act of communication achieves cognitive effects that are large enough to make it worth processing. The novelty is in the second clause. The original formulation required that the cognitive effects could not be achieved by another, more relevant act of communication. The revised clause weakens this demand; it only requires that the cognitive effects cannot be achieved by a more relevant act, *that would be compatible with the communicator's abilities and preferences.*

This revised formulation helps to make one important step towards explaining our findings, but also creates a serious difficulty. Before we discuss these issues, we illustrate the workings of the revised formulation through Sperber and

comparable results to our face-management manipulation. We leave that discovery to future research.

Wilson's own example (p. 275): Peter and Mary are planning a trip to southern France, and Peter is suggesting they could go and meet their old friend Gérard, if that would not take them too far from their way. Peter inquires about the whereabouts of Gérard: 'Where does Gérard lives?' To which Mary replies: 'Somewhere in the south of France.'

Does Mary's reply implicate that she does not know where exactly Gérard lives? Such a claim would follow from the original formulation of the presumption of optimal relevance: If Mary knew that Gérard lived, say, in Toulouse, her vague reply would not have been optimally relevant. But the revised formulation allows another analysis. If Mary is dead against seeing Gérard, then her vague reply is optimally relevant even if she knew that Gérard lived in Toulouse; because the more relevant reply 'Gérard lives in Toulouse' would go against her preferences.

Consider now the utterance 'Some people hated your poem.' Can this utterance convey *everyone hated your poem* in an optimally relevant way? According to the original formulation of optimal relevance, the answer is No: Although the cognitive effects of this interpretation are most likely worth the processing efforts, the same effects could be achieved for less effort by simply saying 'Everyone hated your poem.' However, the revised formulation of optimal relevance takes into account that this simpler formulation would go against the preferences of the speaker, in this case the plausible preference for being gentle rather than inconsiderate. In turn, this permits interpreting 'Some people hated your poem' as *everyone hated your poem* to meet expectations of optimal relevance.

This is an important step in order to account for our findings, but it also raises a difficult problem for the theory. Whilst there is an argument that the utterance 'Some people hated your poem' optimally conveys *everyone hated your poem,* there is a parallel argument that it optimally conveys *some (but not all) hated your poem.* Indeed, as per the standard analysis of 'some,' the cognitive effects of *some (but not all) hated your poem* are worth the processing efforts; and there is no simpler way to convey *some (but not all) hated your poem* than 'Some hated your poem.'

Since saying 'Some people hated your poem' can be optimally relevant to convey either *some (but not all) hated your poem* and *everyone hated your poem,* how do listeners settle on one communicative intention or the other? Relevance Theory assumes that listeners follow a path of least effort, and stop when an interpretation meets their expectations of relevance. In particular, the theory is adamant that listeners never continue looking for another, more effort demanding interpretation, when they have reached an interpretation that meets their expectations of optimal relevance (Sperber & Wilson, 1986/1995). Because both interpretations of 'Some hated your poem' meet expectations of optimal relevance, we arrive at the conclusion that listeners will settle on the interpretation that requires the least processing effort. Now, generally speaking, it is hardly disputable that *some (but not all) X-ed* is an easier interpretation of 'some *X*-ed' than *everyone X-ed;*[5] but this is not the interpretation that participants appear to consider when *X*-ing is a threat to the face of the listener.

At this point, the only way we see for Relevance Theory to account for our results is to make an ad hoc assumption about processing effort. This assumption would be that *everyone X-ed* is usually a costly interpretation of 'some *X*-ed,' but that, for some reason, it gets cheaper when *X*-ing is a threat to the face of the listener; and even cheaper than *some (but not all) X-ed.* Note that we do not currently know what that reason might be. This ad hoc solution has readily testable consequences in terms of response time. Indeed, it predicts that interpreting 'Some *X*-ed' as *some (but not all) X-ed* will take longer than interpreting it as *everyone X-ed* when *X*-ing is a threat to the face of the listener; but that the opposite is true when *X*-ing is a boost to the face of the listener.[6] Of course Relevance theorists may adopt a different explanatory strategy to the one we have outlined. Nevertheless, a test of our prediction would make clear how useful to the theory our explanatory strategy is likely to be.

Conclusion

The findings we have reported raise new theoretical challenges for pragmatic accounts of scalar implicatures, but they also open a new avenue of empirical research into scalars.[7] Researchers now have at their disposal an easy way to manipulate whether a scalar inference is contextually appropriate, in order to observe the effect of this manipulation in reading time or response time studies. In parallel with chronometric issues, the question is opened whether mental workload moderates the effect of face-threatening contexts, and in which direction (De Neys & Schaeken, 2007). If mental workload were to disrupt the effect of face-threatening contexts, the question would be raised whether the processing of face-threatening contexts is correlated with cognitive ability (Feeney et al., 2004). Finally, an important task that was not within the scope of the present paper, but which deserves careful attention, is to address the developmental issue of whether the sensitivity to face-threatening contexts comes to children simultaneously with their understanding of the

[5] Prior experimental data that we reviewed in the introduction established that *some (and possibly all) X-ed* is an easier interpretation of 'some *X*-ed' than *some (but not all) X-ed*. The assumption that we qualify as hardly disputable is that the interpretation of 'some' as *everyone* is more difficult than either the *some (but not all)* or *some (and possibly all)* interpretations.

[6] The generalized account predicts that the interpretation *everyone X-ed* should take longer than the default interpretation *some (but not all) X-ed,* independently of whether *X*-ing is a threat or a boost to the face of the listener.

[7] We note that this future empirical research is likely to use different tasks and measures than those used in the present article. More precisely, it is likely to use reading and response time measures, and the truth or falsity judgement task, which are standard in experiments about the processing of scalar terms. The present research, in contrast, used tasks that were geared at establishing the effectiveness of our contextual manipulation; and it accordingly gave priority to metalinguistic evaluations of word choice as a function of context.

scalar implicature, or whether it occurs at a later developmental stage. Overall, the investigation of face-threatening contexts has the potential to give rise to a new body of empirical data, which will pose new constraints on the way we theorize about scalars.

References

Bezuidenhout, A., & Cutting, J. C. (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, *34*, 433–456.

Bonnefon, J. F., & Villejoubert, G. (2005). Communicating likelihood and managing face: Can we say it is probable when we know it to be certain? In B. G. Bara, L. Barsalou, & M. Buciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 316–321). Mahwah, NJ: Erlbaum.

Bonnefon, J. F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, *17*, 747–751.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Cognition*, *51*, 437–457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*, 434–463.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press. (Original work published 1978)

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In J. Busemeyer, D. L. Medin, & R. Hastie (Eds.), *Decision making from a cognitive perspective* (pp. 275–318). San Diego, CA: Academic Press.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam: John Benjamins.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures.* (Vol. 3, pp. 39–103). Oxford: Oxford University Press.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load –Dual task impact on scalar implicatures. *Experimental Psychology*, *54*, 128–133.

Demeure, V., Bonnefon, J. F., & Raufaste, E. (2008). Utilitarian relevance and face management in the interpretation of ambiguous question/request statements. *Memory and Cognition*, *36*, 873–881.

Demeure, V., Bonnefon, J. F., & Raufaste, E. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *35*, 260–266.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of *some:* Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, *58*, 121–132.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: MIT Press.

Haugh, M. B. (2003). Anticipated versus inferred politeness. *Multilingua*, *22*, 397–413.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context* (pp. 11–42). Washington, DC: Georgetown University Press.

Huang, Y. T., & Snedeker, J. (in press). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*.

Karelitz, T. M., & Budescu, D. V. (2004). You say "probable" and I say "likely": Improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied*, *10*, 25–41.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Noveck, I. A. (2001). When children are more logical than adults: Investigation of scalar implicature. *Cognition*, *78*, 165–188.

Noveck, I. A., & Posada, A. (2003). Characterising the time course of an implicature. *Brain and Language*, *85*, 203–210.

Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences*, *11*, 425–431.

Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of 'scalar inferences'. In N. Burton-Roberts (Ed.), *Pragmatics.* Basingstoke, UK: Palgrave Macmillan.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics/pragmatics interface. *Cognition*, *86*, 253–282.

Sperber, D., & Wilson, D. (1995). *Relevance, communication and cognition*. Oxford: Blackwell. (Original work published 1986)

Williamson, T. (1994). *Vagueness*. London: Routledge.

Youmans, M. (2001). Cross-cultural differences in polite epistemic modal use in American English. *Journal of Multilingual and Multicultural Development*, *22*, 57–73.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353.